REMARKS

Claims 1-24 are pending in the application. Claims 22-24 have been added as new claims. Claims 4, 8, 9, and 13-21 have been cancelled. Claims 1, 2, 3, 5, 6, 10, 11, and 12 have been amended herein. Favorable reconsideration of the application, as amended, is respectfully requested.

I. REJECTION OF CLAIMS UNDER 35 USC § 102

Claims 1-21 stand rejected under 35 USC 102(e) based on being anticipated by US Patent 6,463,415 to St John.

In general the present invention relates to speaker identification and more particularly to a method for performing speaker verification determine whether a speaker uttering a predetermined pass phrase is a registered account holder who has previously provided a voice sample of the same predetermined pass phrase.

St. John deals primarily with automated emotion detection in a speaker's voice (See generally columns 4 through 23), secondarily with automated emotion detection to enhance automated recognition of the spoken words (See generally columns 24 through 29, L30), thirdly with use of automated emotion detection for managing voice messaging (See generally columns 29, L30 through 37, L32), and fourthly with speaker identification (beginning at Column 37, Line 33).

With respect to automated emotion detection, St John teaches that features can be extracted from a digitized voice sample and that the speaker's emotional state can be detected by determining whether the extracted features compare to statistic data compiled from speech features of other speakers who provided voice samples while in a known emotional state.

Automated emotion detection does NOT involve comparing characteristic of a speaker's voice sample (uttering a known phrase) to characteristics of a known speaker uttering the same phrase to determine whether the speaker is the known speaker.

Similarly, recognition of words spoken by a speaker does NOT involve comparing characteristic of a speaker's voice sample (uttering a known phrase) to characters of a known speaker uttering the same phrase. To the contrary, speech recognition exists for the purpose of determining the uttered words when the utterance of the speaker is an unknown sentence or phrase.

The portions of St. John which relate to speaker identification teach comparison of voice characteristics of a speaker uttering a predetermined phrase with a voice characteristic pattern stored in a speaker recognition information storage unit. However, St. John does not teach or suggest the novel systems and methods for performing such comparison as claimed by the applicant.

More specifically St. John teaches comparison of "test voice characteristic information" or "voice input" against the "standard voice characteristic pattern" of the registered speaker to determine if they substantially match" at each of C44, L45-48; C45, L15-18; and C45, L64-67.

However, St. John does not teach or suggest how such comparison is performed and more specifically does not teach or suggest that such comparison could be performed using the novel systems and methods as set forth in the applicant's amended claims.

Claim 1:

Claim 1, as amended, defines a method of performing speaker verification to determine whether a speaker is a registered speaker. The method comprises obtaining an array of frames of compressed audio formants representing the speaker uttering a predetermined pass phrase (P12, L23-L25).

Each frame within the array includes energy data and pitch data characterizing the residue of the speaker uttering the predetermined pass phrase; and formant coefficients characterizing the resonance of the speaker uttering the predetermined pass phrase (P13, L5-L7).

A time domain normalization is performed to normalize the array of compressed audio formants to a sample array of compressed audio formants such that such that the two arrays are of an equal quantity of frames (P13, L14-L20).

The process of determining whether the speaker is the registered speaker comprises: A) generating an array of discrepancy values, each discrepancy value representing the difference between one of: i) an energy value; ii) a pitch value; and iii) a formant coefficient value of a frame of the array and a corresponding energy value; ii) pitch value; and iii) formant coefficient value of a corresponding frame in the sample array (P14, L18-L22); and B) determining whether the array of discrepancy values is within a predetermined threshold (P14,28-P15,L3).

Neither St. John nor the other art of record teaches or suggest such a method for performing speaker verification wherein: A) a time domain normalization is performed such that the two arrays are of an equal quantity of frames; B) generating an array of discrepancy values wherein each discrepancy value represents the difference between one of: i) an energy value; ii) a pitch value; and iii) a formant coefficient value of a frame of the array and a corresponding energy value; ii) pitch value; and iii) formant coefficient value of a corresponding frame in the sample array; nor C) determining whether the array of discrepancy values is within a predetermined threshold.

New Claim 22

New claim 22 is dependent upon claim 1 and is therefore distinguishable over St. John and the other art of record for at least the same reasons.

Further, the method of performing a time domain normalization as described in new claim 22 further distinguishes new claim 22 over St. John and the other art of record.

More specifically, the further defined method comprises comparing the quantity of frames in the array with the quantity of frames in the sample array to determine the quantity of frames to be decimated from the larger of the two arrays such that the two arrays are of an equal quantity for frames (P14,L3-L6).

A pitch decimation group of frames is selected from the larger of the two arrays. The pitch decimation group is the selection of frames which, if decimated yields the best alignment between the pitch values of the two arrays after decimation (P13L29-P14,L2).

An energy decimation group of frames is selected from the larger of the two arrays. The energy decimation group is the selection of frames which, if decimated yields the best alignment between the energy values of the two arrays after decimation (P13L29-P14,L2).

A plurality of formant coefficient decimation groups are selected from the larger of the two arrays. Each formant coefficient decimation group is the selection of frames which, if decimated yields the best alignment between the formant coefficient of the two arrays after decimation (P14L8-L11).

A decimation group of frames is selected from the larger of the two arrays. The decimation group is a quantity of frames equal to the quantity of frames to be decimated and is the a group of frames selected by weighted average from the pitch decimation group, the energy decimation group, and each formant coefficient decimation group. The decimation group of frames is then decimated from the larger of the two arrays such that the two arrays are equal in length.

Claims 2 and 3

Claims 2 and 3 each depend from new claim 22 and can be distinguished over St. John and the other art of record for at least the same reasons. Further, the additional elements and or steps recited in such claims further distinguish such claims over St. John and the other art of record.

Claim 5

Claim 5, as amended, is directed to a method of determining whether a speaker is a registered speaker. The method comprises obtaining compressed audio formants for each frame of an array of frames representing the speaker uttering a predetermined pass phrase.

A time domain normalization is performed to normalize the array to a sample array of frames stored in a memory and representing the registered speaker uttering the predetermined pass phrase. The time domain normalization decimates a portion of the frames of the larger of the two arrays such that the two arrays, after decimation, are of an equal quantity of frames.

The portion of the frames to be decimated is selected by: i) selecting a plurality of audio formant decimation groups, each audio formant decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between a formant coefficient value of each frame of the array and the corresponding formant coefficient value of each frame of the sample array after decimation; and ii) determining a decimation group of frames from the larger of the two arrays, the decimation group being a quantity of frames equal to the quantity of frames to be decimated and being the frames which are selected by weighted average from each of the audio formant decimation groups.

An array of discrepancy values is generated. Each discrepancy value represents the difference between one of an audio formant value of a frame of the array and a corresponding audio formant value of a corresponding frame of the sample array.

It is determined that the remote speaker is the registered speaker if the array of discrepancy values is within a predetermined threshold.

Neither St. John nor the other art of record teaches or suggest such a the applicants claimed method for performing a time domain normalization nor comparing the normalized arrays using a discrepancy value matrix.

Claim 23, 5, and 7

Each of the new claim 23 and claims 5 and 7 each depend from claim 5 and can be distinguished over St. John and the other art of record for at least the same reasons. Further, the additional elements and or steps recited in such claims further distinguish such claims over St. John and the other art of record.

Claim 10

Claim 10, as amended, is directed to a speaker verification server for determining whether a remote speaker is a registered speaker. The server comprises a network interface for receiving, via a packet switched network, compressed audio formants for each frame of an array of frames representing the remote speaker uttering a predetermined pass phrase as audio input to a remote telephony client.

A database stores compressed audio formants for each frame of a sample array of frames representing the registered speaker uttering the predetermined pass phrase as audio input.

A verification application is operatively coupled to each of the network interface and the database for comparing the compressed audio formants of the array of frames to the compressed audio formants of the sample array of frames to determine whether the remote speaker is the registered speaker by: i) performing a time domain normalization of the array to the sample array such that such that the two arrays are of an equal quantity of frames; ii) generating an array of discrepancy values, each discrepancy value representing the difference between one of an audio formant value of a frame of the array and a corresponding audio formant value of a corresponding frame of the sample array; and iii) determining that the remote speaker is the registered speaker if the array of discrepancy values is within a predetermined threshold.

Neither St. John nor the other art of record teaches or suggest such a the applicants claimed method for performing a time domain normalization nor comparing the normalized arrays using a discrepancy value matrix.

Claim 24, 11, and 12

Each of the new claim 24 and claims 11 and 12 each depend from claim 10 and can be distinguished over St. John and the other art of record for at least the same reasons. Further, the additional elements and or steps recited in such claims further distinguish such claims over St. John and the other art of record.

II. CONCLUSION

Accordingly, claims 1 -7, 10-12, and 22-24 are believed to be allowable and the application is believed to be in condition for allowance. A prompt action to such end is earnestly solicited.

Should the Examiner feel that a telephone interview would be helpful to facilitate favorable prosecution of the above-identified application, the Examiner is invited to contact the undersigned at the telephone number provided below.

Should a petition for an extension of time be necessary for the timely reply to the outstanding Office Action (or if such a petition has been made and an additional extension is necessary), petition is hereby made and the Commissioner is authorized to charge any fees (including additional claim fees) to Deposit Account No. 501825.

Respectfully submitted,

Timothy P. O'Hagan
Reg. No. 39,319

DATE: 11-11-04

Timothy P. O'Hagan
8710 Kilkenny Ct
Fort Myers, FL 33912
(239) 561-2300